

SELECTING A PROFILE MODEL FOR USE IN OPTICAL METROLOGY USING A MACHINE LEARNING SYSTEM

BACKGROUND

1. Field of the Invention

[0001] The present application relates to metrology of structures formed on semiconductor wafers, and more particularly to selecting a profile model for use in optical metrology using a machine learning system.

2. Related Art

[0002] Optical metrology involves directing an incident beam at a structure, measuring the resulting diffracted beam, and analyzing the diffracted beam to determine a feature of the structure. In semiconductor manufacturing, optical metrology is typically used for quality assurance. For example, after fabricating a periodic grating in proximity to a semiconductor chip on a semiconductor wafer, an optical metrology system is used to determine the profile of the periodic grating. By determining the profile of the periodic grating, the quality of the fabrication process utilized to form the periodic grating, and by extension the semiconductor chip proximate the periodic grating, can be evaluated.

[0003] One conventional optical metrology system uses a diffraction modeling technique, such as rigorous coupled wave analysis (RCWA), to analyze the diffracted beam. More particularly, in the diffraction modeling technique, a model diffraction signal is calculated based, in part, on solving Maxwell's equations. Calculating the model diffraction signal involves performing a large number of complex calculations, which can be time consuming and costly.

SUMMARY

[0004] In one exemplary embodiment, a profile model can be selected for use in examining a structure formed on a semiconductor wafer using optical metrology by obtaining an initial profile model having a set of profile parameters. A machine learning

system is trained using the initial profile model. A simulated diffraction signal is generated for an optimized profile model using the trained machine learning system, where the optimized profile model has a set of profile parameters with the same or fewer profile parameters than the initial profile model. A determination is made as to whether the one or more termination criteria are met. If the one or more termination criteria are met, the optimized profile model is modified and another simulated diffraction signal is generated using the same trained machine learning system.

DESCRIPTION OF DRAWING FIGURES

[0005] The present invention can be best understood by reference to the following description taken in conjunction with the accompanying drawing figures, in which like parts may be referred to by like numerals:

[0006] Fig. 1 depicts an exemplary optical metrology system;

[0007] Figs. 2A-2E depict exemplary profile models;

[0008] Fig. 3 depicts an exemplary process of selecting a profile model;

[0009] Fig. 4 depicts an exemplary neural network;

[0010] Fig. 5 depicts an exemplary process of training a machine learning system;

[0011] Fig. 6 depicts an exemplary process of testing a machine learning system;

[0012] Fig. 7 depicts another exemplary process of testing a machine learning system; and

[0013] Fig. 8 depicts two exemplary profile models.

DETAILED DESCRIPTION

[0014] The following description sets forth numerous specific configurations, parameters, and the like. It should be recognized, however, that such description is not

intended as a limitation on the scope of the present invention, but is instead provided as a description of exemplary embodiments.

1. Optical metrology

[0015] With reference to Fig. 1, an optical metrology system 100 can be used to examine and analyze a structure on a semiconductor wafer. For example, optical metrology system 100 can be used to determine a feature of a periodic grating 102 formed on wafer 104. As described earlier, periodic grating 102 can be formed in test areas on wafer 104, such as adjacent to a device formed on wafer 104. Alternatively, periodic grating 102 can be formed in an area of the device that does not interfere with the operation of the device or along scribe lines on wafer 104. Furthermore, in some applications, the device can be measured directly.

[0016] As depicted in Fig. 1, optical metrology system 100 can include an optical metrology device with a source 106 and a detector 112. Periodic grating 102 is illuminated by an incident beam 108 from source 106. In the present exemplary embodiment, incident beam 108 is directed onto periodic grating 102 at an angle of incidence θ_i with respect to normal \vec{n} of periodic grating 102 and an azimuth angle Φ (i.e., the angle between the plane of incidence beam 108 and the direction of the periodicity of periodic grating 102). Diffracted beam 110 leaves at an angle of θ_d with respect to normal \vec{n} and is received by detector 112. Detector 112 converts the diffracted beam 110 into a measured diffraction signal, which can include reflectance, $\tan(\Psi)$, $\cos(\Delta)$, Fourier coefficients, and the like.

[0017] Optical metrology system 100 also includes a processing module 114 configured to receive the measured diffraction signal and analyze the measured diffraction signal. As described below, a feature of periodic grating 102 can then be determined using a library-based process or a regression-based process. Additionally, other linear or non-linear profile model extraction techniques are contemplated.

2. Library-based Process

[0018] In a library-based process, the measured diffraction signal is compared to a library of simulated diffraction signals. More specifically, each simulated diffraction signal in the library is associated with a profile model of the structure. When a match is made between the measured diffraction signal and one of the simulated diffraction signals in the library or when the difference of the measured diffraction signal and one of the simulated diffraction signals in the library is within a preset or matching criterion, the profile model associated with the matching simulated diffraction signal in the library is presumed to represent the actual profile of the structure. A feature of the structure can then be determined based on the profile model associated with the matching simulated diffraction signal.

[0019] Thus, with reference again to Fig. 1, in one exemplary embodiment, after obtaining a measured diffraction signal, processing module 114 compares the measured diffraction signal to simulated diffraction signals stored in a library 116. Each simulated diffraction signal in library 116 is associated with a profile model. When a match is made between the measured diffraction signal and one of the simulated diffraction signals in library 116, the profile model associated with the matching simulated diffraction signal in library 116 can be presumed to represent the actual profile of periodic grating 102.

[0020] The set of profile models stored in library 116 can be generated by characterizing a profile model using a set of profile parameters, then varying the set of profile parameters to generate profile models of varying shapes and dimensions. The process of characterizing a profile model using a set of profile parameters can be referred to as parameterizing.

[0021] For example, as depicted in Fig. 2A, assume that profile model 200 can be characterized by profile parameters h_1 and w_1 that define its height and width, respectively. As depicted in Figs. 2B to 2E, additional shapes and features of profile model 200 can be characterized by increasing the number of profile parameters. For example, as depicted in Fig. 2B, profile model 200 can be characterized by profile parameters h_1 , w_1 , and w_2 that define its height, bottom width, and top width,

respectively. Note that the width of profile model 200 can be referred to as the critical dimension (CD). For example, in Fig. 2B, profile parameter w1 and w2 can be described as defining the bottom CD and top CD, respectively, of profile model 200. It should be recognized that various types of profile parameters can be used to characterize profile model 200, including angle of incident (AOI), pitch, n & k, hardware parameters (e.g., polarizer angle), and the like.

[0022] As described above, the set of profile models stored in library 116 (Fig. 1) can be generated by varying the profile parameters that characterize the profile model. For example, with reference to Fig. 2B, by varying profile parameters h1, w1, and w2, profile models of varying shapes and dimensions can be generated. Note that one, two, or all three profile parameters can be varied relative to one another.

[0023] Thus, the profile parameters of the profile model associated with a matching simulated diffraction signal can be used to determine a feature of the structure being examined. For example, a profile parameter of the profile model corresponding to a bottom CD can be used to determine the bottom CD of the structure being examined.

[0024] With reference again to Fig. 1, the number of profile models and corresponding simulated diffraction signals in the set of profile models and simulated diffraction signals stored in library 116 (i.e., the resolution and/or range of library 116) depends, in part, on the range over which the set of profile parameters and the increment at which the set of profile parameters are varied. In one exemplary embodiment, the profile models and the simulated diffraction signals stored in library 116 are generated prior to obtaining a measured diffraction signal from an actual structure. Thus, the range and increment (i.e., the range and resolution) used in generating library 116 can be selected based on familiarity with the fabrication process for a structure and what the range of variance is likely to be. The range and/or resolution of library 116 can also be selected based on empirical measures, such as measurements using atomic force microscopy (AFM), scanning electron microscopy (SEM), and the like.

[0025] For a more detailed description of a library-based process, see U.S. Patent Application Ser. No. 09/907,488, titled GENERATION OF A LIBRARY OF PERIODIC

GRATING DIFFRACTION SIGNALS, filed on July 16, 2001, which is incorporated herein by reference in its entirety.

3. Regression-based Process

[0026] In a regression-based process, the measured diffraction signal is compared to a simulated diffraction signal generated prior to the comparison (i.e., a trial simulated diffraction signal) using a set of profile parameters (i.e., trial profile parameters) for a profile model. If the measured diffraction signal and the trial simulated diffraction signal do not match or when the difference of the measured diffraction signal and the trial simulated diffraction signal is not within a preset or matching criterion, another trial simulated diffraction signal is generated using another set of profile parameters for another profile model, then the measured diffraction signal and the newly generated trial simulated diffraction signal are compared. When the measured diffraction signal and the trial simulated diffraction signal match or when the difference of the measured diffraction signal and the trial simulated diffraction signals is within a preset or matching criterion, the profile model associated with the matching trial simulated diffraction signal is presumed to represent the actual profile of the structure. The profile model associated with the matching trial simulated diffraction signal can then be used to determine a feature of the structure being examined.

[0027] Thus, with reference again to Fig. 1, in one exemplary embodiment, processing module 114 can generate a trial simulated diffraction signal for a profile model, and then compare the measured diffraction signal to the trial simulated diffraction signal. As described above, if the measured diffraction signal and the trial simulated diffraction signal do not match or when the difference of the measured diffraction signal and the trial simulated diffraction signals is not within a preset or matching criterion, then processing module 114 can iteratively generate another trial simulated diffraction signal for another profile model. In one exemplary embodiment, the subsequently generated trial simulated diffraction signal can be generated using an optimization algorithm, such as global optimization techniques, which includes simulated annealing, and local optimization techniques, which includes steepest descent algorithm.

[0028] In one exemplary embodiment, the trial simulated diffraction signals and profile models can be stored in a library 116 (i.e., a dynamic library). The trial simulated diffraction signals and profile models stored in library 116 can then be subsequently used in matching the measured diffraction signal. Alternatively, library 116 can be omitted from optical metrology system 100.

[0029] For a more detailed description of a regression-based process, see U.S. Patent Application Ser. No. 09/923,578, titled METHOD AND SYSTEM OF DYNAMIC LEARNING THROUGH A REGRESSION-BASED LIBRARY GENERATION PROCESS, filed on August 6, 2001, which is incorporated herein by reference in its entirety.

4. Selecting Optimal Profile Model

[0030] The accuracy, complexity, and length of time needed to perform a library-based process and/or regression-based process depends, in part, on the complexity of the profile model used. For example, increasing the complexity of the profile model by adding a profile parameter can increase accuracy. However, the increased complexity of the profile model can increase the complexity and the amount of time needed to perform the library-based process and/or regression-based process. Thus, with reference to Fig. 3, in one exemplary embodiment, an optimal profile model to be used in a library-based process and/or regression-based process is selected using exemplary process 300.

[0031] In step 302, a measured diffraction signal is obtained. In the present exemplary embodiment, the measured diffraction signal from a structure to be examined is obtained using an optical metrology device, such as a reflectometer, ellipsometer, and the like. Note that the structure used to obtain the measured diffraction signal can be the actual structure to be examined or a representative structure of the actual structure to be examined.

[0032] In step 304, an initial profile model is obtained. The initial profile model has a set of profile parameters that characterize the structure to be examined. In the present exemplary embodiment, the initial profile model is the most complex profile model that will be used in process 300, and eventually the library-based process and/or regression-

based process. Thus, if process 300 is iterated, progressively simpler profile models are used in iterating process 300. For example, the initial profile model used in the first iteration of process 300 can include six profile parameters. If process 300 is iterated, the profile model used in the second iteration of process can be simplified to include five profile parameters. In the present exemplary embodiment, the initial profile model can be selected by a user or can be automatically selected using a default profile model.

[0033] In step 306, a machine learning system is trained using the initial profile model. With reference to Fig. 1, in one exemplary embodiment, the machine learning system employs a machine learning algorithm, such as back-propagation, radial basis function, support vector, kernel regression, and the like. For a more detailed description of machine learning systems and algorithms, see “Neural Networks” by Simon Haykin, Prentice Hall, 1999, which is incorporated herein by reference in its entirety. See also U.S. Patent Application Ser No. 10/608,300, titled OPTICAL METROLOGY OF STRUCTURES FORMED ON SEMICONDUCTOR WAFERS USING MACHINE LEARNING SYSTEMS, filed on June 27, 2003, which is incorporated herein by reference in its entirety.

[0034] With reference to Fig. 4, in one exemplary implementation, the machine learning system is a neural network 400 using a back-propagation algorithm. Neural network 400 includes an input layer 402, an output layer 404, and a hidden layer 406 between input layer 402 and output layer 404. Input layer 402 and hidden layer 406 are connected using links 408. Hidden layer 406 and output layer 404 are connected using links 410. It should be recognized, however, that neural network 400 can include any number of layers connected in various configurations.

[0035] As depicted in Fig. 4, input layer 402 includes one or more input nodes 412. In the present exemplary implementation, an input node 412 in input layer 402 corresponds to a profile parameter of the profile model that is inputted into neural network 400. Thus, the number of input nodes 412 corresponds to the number of profile parameters used to characterize the profile model. For example, if a profile model is characterized using two profile parameters (e.g., top and bottom critical dimensions),

input layer 402 includes two input nodes 412, where a first input node 412 corresponds to a first profile parameter (e.g., a top critical dimension) and a second input node 412 corresponds to a second profile parameter (e.g., a bottom critical dimension).

[0036] In neural network 400, output layer 404 includes one or more output nodes 414. In the present exemplary implementation, each output node 414 is a linear function. It should be recognized, however, that each output node 414 can be various types of functions. Additionally, in the present exemplary implementation, an output node 414 in output layer 404 corresponds to a dimension of the simulated diffraction signal that is outputted from neural network 400. Thus, the number of output nodes 414 corresponds to the number of dimensions used to characterize the simulated diffraction signal. For example, if a simulated diffraction signal is characterized using five dimensions corresponding to, for example, five different wavelengths, output layer 404 includes five output nodes 414, wherein a first output node 414 corresponds to a first dimension (e.g., a first wavelength), a second output node 414 corresponds to a second dimension (e.g., a second wavelength), etc. Additionally, for increased performance, neural network 400 can be separated into a plurality of sub networks based on separate components of the simulated diffraction signal and/or dimensions of the components of the simulated diffraction signal.

[0037] In neural network 400, hidden layer 406 includes one or more hidden nodes 416. In the present exemplary implementation, each hidden node 416 is a sigmoidal transfer function or a radial basis function. It should be recognized, however, that each hidden node 416 can be various types of functions. Additionally, in the present exemplary implementation, the number of hidden nodes 416 is determined based on the number of output nodes 414. More particularly, the number of hidden nodes 416 (m) is related to the number of output nodes 414 (n) by a predetermined ratio ($r = m/n$). For example, when $r = 10$, there are 10 hidden nodes 416 for each output node 414. It should be recognized, however, that the predetermined ratio can be a ratio of the number of output nodes 414 to the number of hidden nodes 416 (i.e., $r = n/m$). Additionally, it should be recognized that the number of hidden nodes 416 in neural network 400 can be adjusted after the initial number of hidden nodes 416 is determined based on the

predetermined ratio. Furthermore, the number of hidden nodes 416 in neural network 400 can be determined based on experience and/or experimentation rather than based on the predetermined ratio.

[0038] With reference to Fig. 5, an exemplary process 500 is depicted for training a machine learning system. In exemplary process 500, the machine learning system is trained using a set of training input data and a set of training output data, where an input data in the set of training input data has a corresponding output data in the set of training output data to form an input and an output data pair.

[0039] In step 502, the set of training input data is obtained. In the present exemplary embodiment, the training input data includes a set of profile models generated based on the initial profile model. More particularly, the set of profile models is generated by varying one or more profile parameters that characterize the initial profile model, either alone or in combination. The one or more profile parameters are varied over one or more ranges based on the expected range of variability in the actual profile of the structure to be examined, the expected range of variability is determined either empirically or through experience. For example, if the actual profile of the structure to be examined is expected to have a bottom critical dimension that can vary between x_1 and x_2 , then the set of profile models used as the training input data can be generated by varying the profile parameter in the initial profile model corresponding to the bottom critical dimension between x_1 and x_2 .

[0040] In step 504, the set of training output data is obtained. In the present exemplary embodiment, the training output data includes a set of diffraction signals. A diffraction signal in the set of diffraction signals used as the training output data corresponds to a profile model in the set of profile models used as the training input data. Each diffraction signal in the set of diffraction signals can be generated based on each profile model in the set of profile models using a modeling technique, such as rigorous coupled wave analysis (RCWA), integral method, Fresnel method, finite analysis, modal analysis, and the like. Alternatively, each diffraction signal in the set of diffraction signals can be generated based on each profile model in the set of profile models using an

empirical technique, such as measuring a diffraction signal using an optical metrology device, such as an ellipsometer, reflectometer, atomic force microscope (AFM), scanning electron microscope (SEM), and the like. Thus, a profile model from the set of profile models and the corresponding diffraction signal from the set of diffraction signals form a profile model/diffraction signal pair. Although there is a one-to-one correspondence between a profile model and a diffraction signal in the profile model/diffraction signal pair, note that there does not need to be a known relation, either analytic or numeric, between the profile model and the diffraction signal in the profile model/diffraction signal pair.

[0041] In step 506, simulated diffraction signals are generated with the machine learning system using the training input data as inputs to the machine learning system. In step 508, a determination is made as to whether one or more termination criteria are met. In the present exemplary embodiment, a termination criterion can be based on an analysis of the diffraction signals (i.e., the diffraction signals in the training output data and the simulated diffraction signals generated by the machine learning system), such as a cost function value, a Goodness-of-Fit (GOF) value, various curve fitting metrics, and the like. Alternatively or additionally, a termination criterion can be based on an analysis of the profile models, such as correlation, sensitivity, confidence interval, and the like. It should be recognized that the determination made in step 508 can be based on a combination of any two or more termination criteria.

[0042] A cost function determined between two diffraction signals is illustrated by the equations below, where V_1 and V_2 are two vectors of size n , and the cost function of V_1 relative to V_2 is:

$$Cost(V_1, V_2) = \left(\sum_{i=1}^n (V_{1i} - V_{2i})^p \right)^{1/p}$$

where i represents the i th member of the vector and p is an arbitrary number associated with the metric. The first vector is the set of signal values for a first diffraction signal,

and the second vector is the corresponding set of signal values for a second diffraction signal.

[0043] A goodness of fit (GOF) is a measure of the proximity of two sets values. For example, when ellipsometric measurements are used, GOF can be determined based on values for $\tan \psi$ and $\cos \Delta$, where $\tan \psi$ and $\cos \Delta$ are represented by a single vector of n dimensions:

$$S = [\tan \psi_1 \tan \psi_2 \dots \tan \psi_{n/2} \cos \Delta_1 \cos \Delta_2 \dots \cos \Delta_n]$$

[0044] One commonly used formula for GOF between a first signal S_1 compared to a second signal S_2 is:

$$GOF = 1 - \frac{\sum_i^n (S_2(i) - S_1(i))^2}{\sum_i^n (S_1(i) - \bar{S}_1)^2}$$

$$\text{where } \bar{S}_1 = \frac{\sum_i^n S_1(i)}{n}$$

where i represents the i *th* point for comparison, n is the total number of points of comparison.

[0045] A correlation coefficient, r , between two profile parameters can be calculated using the formula:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

where x_i and y_i is a pair of profile parameters, \bar{x} is the mean of x_i 's and \bar{y} is the mean of y_i 's. The value of r lies between -1 and +1 inclusive. A correlation coefficient value of +1 can correspond to complete positive correlation and a value of -1 can correspond to complete negative correlation. A value of r close to zero can correspond to the x and y profile parameters not being correlated.

[0046] A sensitivity of a diffraction signal to changes in one or more profile parameters can be determined by changing one profile parameter by a small amount and keeping the other profile parameters constant. For example, the sensitivity of profile parameter x0 may be tested by adding one nanometer to the nominal value while keeping profile parameters x1, x2, and x3 at nominal value. If there is no noticeable change in the diffraction signal (x0 at nominal plus 1 nm), then x0 has low sensitivity. The other profile parameters can similarly be changed while holding the rest constant in order to test the sensitivity of each profile parameter.

[0047] The sensitivity of a profile parameter may be quantitatively expressed by calculating the sum-square-error (SSE) of the changed diffraction signal compared to the diffraction signal using nominal values. The SSE formula is as follows:

$$SSE = \sum_{i=1}^n (S_o(i) - S_1(i))^2$$

where i is the signal point, typically at a preset wavelength, n is the number of signal points, S₀ is the diffraction signal value using nominal values of profile parameters, S₁ is the diffraction signal value using nominal plus change in one of the profile parameters.

[0048] A confidence interval of a profile parameter can be determined by the amount of change from a nominal value of the profile parameter, where the change in the diffraction signals is greater than the noise level. The noise in the diffraction signals may be due to system noise, for example, noise from the measurement devices, or the noise may be simulated. The confidence interval is generally expressed as a multiple of the standard deviation sigma, σ, of the profile parameter. The standard deviation for a profile parameter can be calculated from measured values of the profile parameter, using the formula:

$$\sigma = \sqrt{\left(\frac{1}{(N-1)} \right) \sum (x_i - x_{av})^2}$$

where N is the number of measurements, x_i is the ith value of the profile parameter x, and x_{av} is the average value of the profile parameter x. In the present exemplary embodiment, a confidence interval of 3 sigmas can be used.

[0049] The confidence interval is typically calculated from a given set of sample input data representing actual measurements off the wafer structure. The confidence interval may also be calculated using simulated random noise introduced in the measurement data for the profile parameter.

[0050] As depicted in Fig. 5, if the one or more termination criteria are not met, step 506 is repeated. In the present exemplary embodiment, before repeating step 506, the machine learning system is adjusted. For example, when the machine learning system is a neural network, the weights used in the functions or the number of hidden nodes of the neural network can be adjusted. After adjusting the machine learning system, step 506 is repeated to generate diffraction signals using the training input data as inputs to the adjusted machine learning system. Alternatively or additionally, a new set of training input and output data can be obtained, and then diffraction signals are generated using the new training input data as inputs to the machine learning system.

[0051] It should be recognized that training process 500 can include the use of an optimization technique, such as gradient descent, linear programming, quadratic programming, simulated annealing, Marquardt-Levenberg algorithm, and the like. Additionally, training process 500 is depicted as batch training, where diffraction signals are generated for all of the profile models in the training input data as a batch. For a more detailed description of batch training, see “Neural Networks” by Simon Haykin, which has been cited above. It should be recognized, however, that a diffraction signal can be generated for each of the profile models in the training input data one at a time.

[0052] Furthermore, training process 500 depicted in Fig. 5 illustrates a back-propagation algorithm. However, it should be recognized that various training algorithms can be used, such as radial basis network, support vector, kernel regression, and the like.

[0053] With reference to Fig. 6, an exemplary process 600 is depicted for testing a machine learning system. In the present exemplary embodiment, after a machine learning system has been trained, the machine learning system can be tested to confirm that it has been properly trained. It should be recognized, however, that this testing process can be omitted in some applications.

[0054] In 602, a set of testing input data is obtained. In 604, a set of testing output data is obtained. In the present exemplary embodiment, the testing input data includes a set of profile models, and the testing output data includes a set of diffraction signals. The set of testing input data and set of testing output data can be obtained using the same process and techniques described above during the training process. The set of testing input data and set of testing output data can be the same as or a subset of the training input data and training output data. Alternatively, the set of testing input data and set of testing out data can be different than the training input data and training output data.

[0055] In 606, simulated diffraction signals are generated with the machine learning system using the testing input data as inputs to the machine learning system. In 608, a determination is made as to whether one or more termination criteria are met. In the present exemplary embodiment, a termination criterion can be based on an analysis of simulated diffraction signals (i.e., the simulated diffraction signals in the training output data and the simulated diffraction signals generated by the machine learning system), such as a cost function value, a Goodness-of-Fit (GOF) value, various curve fitting metrics, and the like. Alternatively or additionally, a termination criterion can be based on an analysis of the profile models, such as correlation, sensitivity, confidence interval, and the like. It should be recognized that the determination made in 608 can be based on a combination of any two or more termination criteria.

[0056] In 610, if the one or more termination criteria are not met, the machine learning system is re-trained. When the machine learning system is re-trained, the machine learning system can be adjusted. For example, when the machine learning system is a neural network, the weights used in the functions or the number of hidden nodes of the neural network can be adjusted. Alternatively or additionally, the selection and number of the training input and output variables can be adjusted.

[0057] With reference to Fig. 7, another exemplary process 700 is depicted for testing or validating a machine learning system. In the present exemplary embodiment, a first machine learning system can be tested or validated by training a second machine learning system.

[0058] In step 702, the second machine learning system is trained using the same set of training data used to train the first machine learning system. However, the training input data used in training the first machine learning system is used as the training output data in training the second machine learning system, and the training output data used in training the first machine learning system is used as the training input data in training the second machine learning system. Thus, when the first machine learning system is trained using profile models as inputs and diffraction signals as outputs, the second machine learning system is trained using diffraction signals as inputs and profile models as outputs.

[0059] After the second machine learning system has been trained, in step 704, one or more profile models are used as inputs to generate one or more simulated diffraction signals using the first machine learning system. In step 706, the one or more simulated diffraction signals generated by the first machine learning system are used as inputs to generate one or more profile models using the second machine learning system.

[0060] In step 708, the one or more profile models generated by the second machine learning system and the one or more profile models that were used as inputs into the first machine learning system can be analyzed. For example, if the difference between the profile models is within an acceptable tolerance, the first machine learning system is validated.

[0061] An empirical risk minimization (ERM) technique can be used to quantify how well the trained machine learning system can generalize to new input. For a more detailed description of ERM, see “Statistical Learning Theory” by Vladimir N. Vapnik, Wiley-Interscience, September 1998, which is incorporated herein by reference in its entirety.

[0062] With reference again to Fig. 3, after the machine learning system has been trained using the initial profile model, in step 308, the machine learning system is used to generate a simulated diffraction signal for an optimized profile model. In the present exemplary embodiment, the optimized profile model has a set of profile parameters with the same or fewer profile parameters than the initial profile parameter. Note that the

optimized profile model can be the same as the initial profile model in the first iteration of process 300.

[0063] In step 310, a determination is made as to whether one or more termination criteria are met. In the present exemplary embodiment, a termination criterion can be based on an analysis of simulated diffraction signals (i.e., the simulated diffraction signals in the training output data and the simulated diffraction signals generated by the machine learning system), such as a cost function value, a Goodness-of-Fit (GOF) value, various curve fitting metrics, and the like. Alternatively or additionally, a termination criterion can be based on an analysis of the profile models, such as correspondence, correlation, sensitivity, confidence interval, and the like. It should be recognized that the determination made in 310 can be based on a combination of any two or more termination criteria.

[0064] In the present exemplary embodiment, when cost function is included as a termination criterion, a cost function value can be determined between the simulated diffraction signal and the measured diffraction signal. The determined cost function can then be compared to a preset cost function value to determine if the determined cost function value is less than or equal to the preset cost function value. The preset cost function value may be set at a specific number, for example, 0.05.

[0065] When GOF value is included as a termination criterion, a GOF value can be determined between the simulated diffraction signal and the measured diffraction signal. The determined GOF value can then be compared to a preset GOF value to determine if the determined GOF value is less than or equal to the preset GOF value. The preset GOF value may be set at a specific number, for example 0.95.

[0066] When correspondence is included as a termination criterion, a correspondence is obtained between the profile parameters of the optimized profile model and the dimensions of the actual profile that corresponds to the measured diffraction signal. In the present exemplary embodiment, the dimensions of the actual profile can be obtained using SEM.

[0067] When correlation is included as a termination criterion, a correlation coefficient can be determined between a pair of profile parameters of the optimized profile model. The determined correlation coefficient can then be compared to a preset correlation coefficient to determine if the determined correlation coefficient is less than or equal to the preset correlation coefficient.

[0068] When parameter sensitivity is included as a termination criterion, a sensitivity can be determined for each profile parameter of the optimized profile model. The determined sensitivity can then be compared to a preset sensitivity to determine if the determined sensitivity is less than or equal to the preset sensitivity coefficient.

[0069] When confidence interval is included as a termination criterion, a confidence interval is determined for each profile parameter of the optimized profile model. The determined confidence interval can then be compared to a preset confidence interval to determine if the determined confidence interval is less than or equal to the preset confidence interval. The preset confidence interval may be set to any number of sigma, such as three-sigma.

[0070] In step 312, if the one or more termination criteria are not met, the optimized profile model is modified and steps 308 and 310 are iterated. In the present exemplary embodiment, the optimized profile model is modified to reduce the number of profile parameters used to characterize the optimized profile model used in iterating step 308.

[0071] For example, with reference to Fig. 8, assume that in a first iteration of step 308 (Fig. 3), optimized profile model 800 was used. As depicted in Fig. 8, optimized profile model 800 is characterized by six profile parameters (i.e., thickness of a first thin film layer (t_1), thickness of a second thin film layer (t_2), thickness of a third thin film layer (t_3), bottom critical dimension (BCD), a top critical dimension (TCD), and a height (h)). In step 312 (Fig. 3), assume that optimized profile model 800 is modified as optimized profile model 802 by eliminating the bottom critical dimension (BCD). Optimized profile model 802 is then used in iterating step 308 (Fig. 3).

[0072] With reference again to Fig. 3, in performing step 312, a user can specify the modification to the optimized profile model. For example, selection of the profile parameter to be eliminated is one way to specify the modification to the optimized profile model. Alternatively, the selection of the profile parameter to be eliminated can be made using one or more selection criteria, such as correlation, sensitivity, confidence interval, and the like.

[0073] For additional examples of profile model selection processes, see U.S. Patent Application Ser. No. 10/206,491, titled MODEL AND PARAMETER SELECTION FOR OPTICAL METROLOGY, filed on July 25, 2002, which is incorporated herein by reference in its entirety. See also U.S. Patent Application No. 10/397,631, titled OPTIMIZED MODEL AND PARAMETER SELECTION FOR OPTICAL METROLOGY, filed on March 25, 2003, which is incorporated herein by reference in its entirety.

[0074] With reference again to Fig. 3, in the present exemplary embodiment, in iterating step 308, the same machine learning system is used. Because the optimized profile model used in generating the simulated diffraction signal in step 308 includes the same or fewer profile parameters than the initial profile model used to train the machine learning system in step 302, the machine learning system does not need to be retrained, which reduces the amount of time to generate the simulated diffraction signal in step 308.

[0075] In the exemplary embodiment, after selecting an optimized profile model using process 300, a profile refinement process can be used to select at least one profile parameter of the optimized profile model and set the at least one profile parameter to a determined value. The at least one profile parameter can be selected using one or more selection criteria, such as correlation, fabrication process knowledge, historical information, the ability to obtain measurements from metrology tools, and the like. The determined value for the at least one profile parameter can be obtained from a variety of sources, such as specific measurements of the at least one profile parameter, profile extraction, theoretical and/or empirical data, estimates based on simulations of fabrication

recipes using semiconductor fabrication simulation systems, mathematical and/or statistical techniques, averaging techniques, and the like.

[0076] For example, assume a selection criteria includes a correlation of at least 0.95 or higher. Now assume that an optimized profile model includes a width parameter and a thickness parameter with a correlation greater than 0.95. Thus, in this example, the width parameter and/or the thickness parameter is selected and set to a determined value.

[0077] Assume that the thickness parameter in the example above is selected. Now assume that the determined value is obtained using an averaging technique. More particularly, in the present example, multiple thickness measurements of the selected thickness parameter on a wafer are obtained. An average thickness measurement of the selected thickness parameter is then calculated from the multiple thickness measurements. The selected thickness parameter is then set to the average thickness measurement.

[0078] Note that when a profile refinement process is used with a machine learning system, a selected profile parameter can be set to any value. However, when a profile process is used with a library-based system, a selected profile parameter is preferably set based on a constraint of the library, such as the resolution of the library. For example, if a profile process is used with a machine learning system and an average thickness measurement is 50.25 nanometers, then the selected thickness parameter can be set to 50.25 nanometers. However, if a profile process is used with a library-based system and the library includes thickness parameters at intervals of 50, 55, and 60 nanometers, then the selected thickness parameter is set to 50 nanometers.

[0079] For a more detailed description of a profile refinement process, see U.S. Patent Application Ser. No. 10/735,212, titled PARAMETRIC OPTIMIZATION OF OPTICAL METROLOGY MODEL, filed on December 12, 2003, which is incorporated herein by reference in its entirety.

[0080] Although exemplary embodiments have been described, various modifications can be made without departing from the spirit and/or scope of the present invention.

Therefore, the present invention should not be construed as being limited to the specific forms shown in the drawings and described above.